The Role of Measurement Error in Modelling and Simulation

Jens Wendelstorf

TU-Clausthal, Germany
jens.wendelstorf@tu-clausthal.de
jens@wendelstorf.de

preprint published as:
Hochschulschrift, 28.01.2021, TU-Clausthal, Germany
DOI: 10.21268/20210125-0

Abstract
Modelling, Simulation and computational Optimization (MSO) becomes more and more ubiquitous in science and technology. Supporting the quantification of all branches of science it is mission critical to address a potential area of unexpressed ignorance:
the measurement error and its counterpart in modelling, the prediction error.
This paper introduces the most important categories of measurement errors and their influence on the predictability of measurements by modelling and simulation. It explains the concept of model fit parameters and their indirect measurement. It counteracts to the unjustified uncertainty specification phobia by falsifying the misleading concept of errorbars for two or more fit-parameters and gives a survey on current and future model uncertainty computations made possible by the ubiquitous computational resources.

1 Introduction to **scientific modelling and simulation**

Reproducibility is one of the paradigmatic principles in science and an indicator for inefficiencies in the scientific enterprise [1]. It can be used to identify true science in the noise of the social activity with the same name. Without knowledge of the uncertainties of a specific measurement or model prediction, it is impossible to assess it' s reproducibility. Additionally, open data publication may be seen as a prerequisite to reproducibility [2].

Further important paradigmatic principles of science are **simplicity** (Occam's razor) and **falsifiability** [3]. Additionally, for us as humans, **disinterest** (no reward) becomes an important issue balancing our prevalent behaviours. Denying the hypothesis of science as a pure social activity, there crystalize three categories of hard tasks in science:

- Discover a simple concept for the description of a large class of real phenomena.
- Discover a simple model or theory quantitatively describing (and predicting) a relevant part of the real world.

- Provide a relevant observation of the (actual and past) real world which is relevant and falsifiable.

These categories have qualitative and quantitative aspects and these interact, e.g. more accurate measurements (quanititative aspect) have often led to new and important conceptual or theoretical discoveries. For all these categories, uncertainty specifications are mission critical.

In highly developed scientific disciplines, modelling by sophisticated computer codes and simulation by the execution of these codes has seen remarkable progress (see e.g. [4]) and the ever increasing resource requirements demand for the consideration of uncertainties in order to foresee the priorities for these resources. In other scientific disciplines, while outgrowing from their infancy, modelling and simulation is of the same importance and uncertainty quantification by using simple and easy to use software tools is one of the key requirements for distinguishing truth from cargo cult.

Modelling, simulation and optimization (**MSO**) activities are often called ***computational X*** where ***X*** is the name of the branch of science. Quality assurance for these activities requires a carefull consideration of those uncertainties relevant to their predictions.

1.1 Can models be validated?

The great tragedy of science - the slaying of a beautiful hypothesis by an ugly fact
[T.H. Huxley'1893]

In order to distinguish **scientific modelling and simulation** from marketing activities, the widely accepted quality assurance activity is called *model validation*. It tries to justify a model by showing its ability to compute the expected results. This activity is important because it can result in a **model falsification** [3], although such a finding is rarely published. Nevertheless, the belief in a validated model may later on become a misbelief, because the range of experimental data used for the validation process may have been too narrow and even a single reproducebale experiment not predictable by the model will falsify it.
Models can not be validated, models can only be falsified!
The term validation has to be used very carefully as an abbreviation for extensive and unsuccessfull attempts to falsify a model by undertaking accurate experiments over the full range of all of the models variables.

1.2 What is a model?

A model is a simplified representation of a system (process). An ideal model requires only a minimum number of input variables for making predictions and these variables should be known or measureable. So one important modelling objective is to

- **derive your models in variables insensitive to measurement errors!**

Scientific models rely on scientific theories and the tracing of all quantities back to accurate laboratory measurements and natural constants. Real world engineering models may contain fit parameters which later eventually become the constants of new correlations or even principles. Sometimes, these fit parameters are not easy to identify or the user of the modelling software is not allowed to change their values.

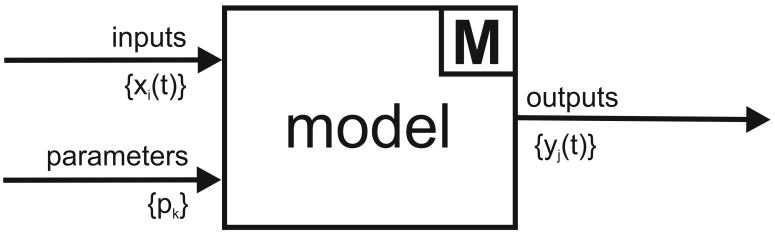


Figure 1: Simplified black box view of a model (after [5,6]).

Modelling and Simulation is called the third pillar of scientific research and an enabling technology in industrial R&D [7]. A computer **model** $\mathcal{M}$ is the simplified translation of knowledge about a process into an algorithm able to perform simulations. As drafted in **figure 1**, in such a simulation, the specified **inputs** (n_i input variables $x_i(t)$ and n_k parameters p_k) are used to compute the **outputs** (n_j results $y_j(t)$). The inputs and outputs are not necessarily lumped or time dependent, they may also be distributed in space or other coordinates (*distributed* or *lumped parameter models*).

In general, the parameters of a model may also be known from independant laboratory measurements or may be computational parameters which need to be determined by simulating cases where the outputs are exactly known. Here, we concentrate on those parameters which have to be determined from n_p measurements of the outputs y_j for known inputs x_i, known as the process specific **fit parameters** p_k.

The applicable scientific branch is called **system identification** or **model identification** [8] because for the prediction of a specific y_j one has to identify the best combination of a model $\mathcal{M}$ and a representative set of fit parameter vectors $\{\vec{p}_m\}$. By considering measurement uncertainties, the cardinal number of this set n_m may become large.

It is the responsibility of the experimentalist to provide and justify quantitative error information for his experiments. The modeller has the responsibility to provide and justify quantitative error information for the predictions of his model. In this paper the basic definitions and steps for this challenging task are illustrated. Obviously, the difficulties of this exercise may originate from the structure of the model itself and can be controlled by a carefull definition and implementation of the model:

- A model must **predict measureable quantities**. If the model has fit parameters, for each of these fit parameters, there have to be measurements of outputs showing a sufficient sensitivity with respect to this fit parameter. Even a single fit parameter with no sensitivity on some of the measureable outputs falsifies the identifiability of the model.

A model is thus not only falsifiable by measurements contradictive to its predictions but may also be inable to predict its outputs with a finite uncertainty. A model may be **unidentifiable**, i.e. it's fit parameters can not be determined. If a modeller provides predictions in areas where no experimental data is known at the time of the prediction, she leaves the area of data science and takes the first step to scientific modelling.

2. The **measurement uncertainty**

2.1 Definition

Before the scientific language incorporated political and psychological considerations, our subject was termed **measurement *error***. As of today, the universal constants are known within specified bounds and accurate and precise measurements are possible in most branches of science (quantification). Thus the term uncertainty is justified. Nevertheless, the falsification principle remains an essential paradigm:
A scientific theory (or prediction) must be falsifiable by measurements and measurements are falsified by the establishment of (more accurate) measurements providing results not within the previously indicated regions of values.

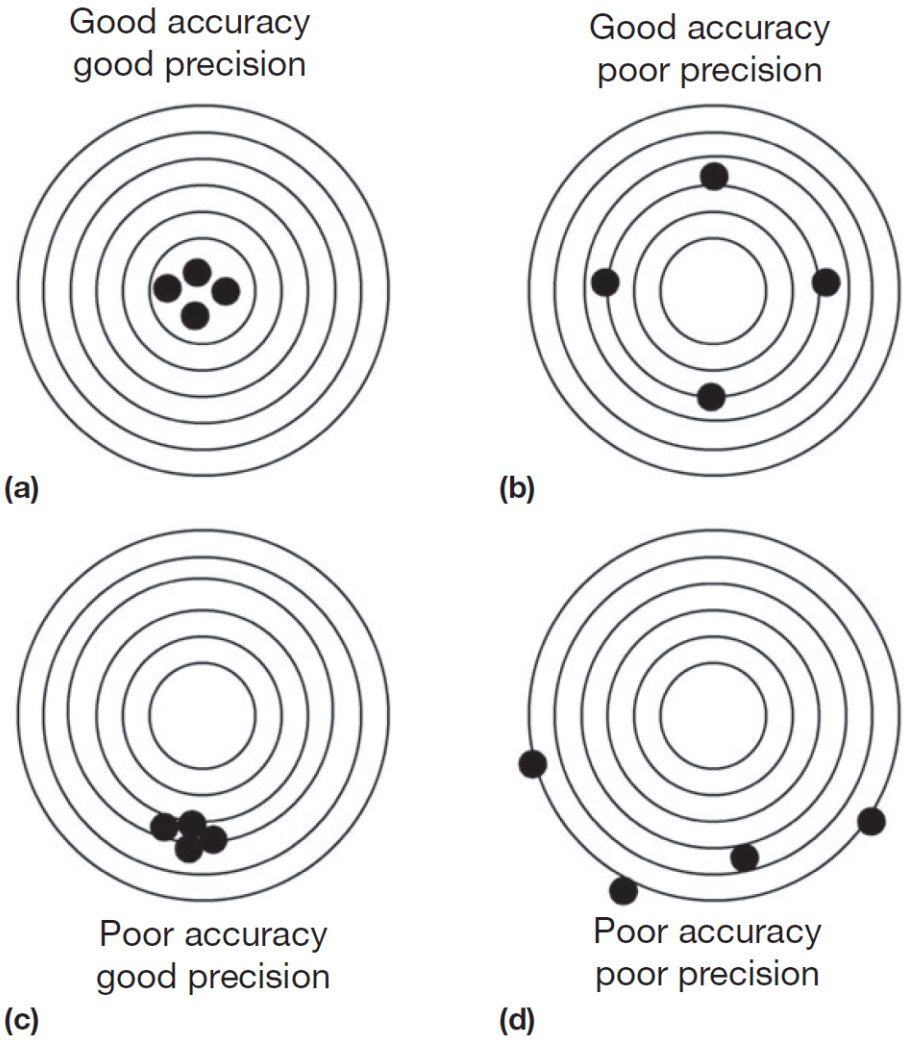


Figure 1: The four extreme combinations of epistemic (accuracy) and alleatoric (precision) uncertainty;
the unknown true value is the bull's eye [9].

The real uncertainty of a measurement is a superposition of two qualitatively and quantitatively different types of uncertainties, as shown in figure 1 and discussed in the following subsections.

2.2 The measurement **precision** describing random uncertainty (noise)

The reproducibility of a measurement (e.g. a radioactive decay experiment) may be determined by the influence of random processes (noise) on the individual result. If such a noise is the only source of uncertainty, the repetition of the measurement provides **statistical information** which can be used to **describe the** so called **precision** of the measurement. For more than 8 repetitions, the standard deviation σ and the mean value μ become increasingly usefull with an increasing number of repetitions. For more than about hundred repetitions, the results begin to show the statistical properties of the measurement. For a gaussian distribution function it is commonly accepted to use the statistical 95% confidence interval for the specification of the random error called measurement precision

$$\in_{\text{prec}} = 2\,\sigma_{\text{random processes}}$$

The most important property of the measurement precision is the possibility to determine it from the measured values themselves by repeating the measurement under constant conditions. For enhancing the standard deviation σ_μ of the mean value by a factor of two, four times more experiments are required. This type of uncertainty is also called **aleatoric uncertainty**, irreducible uncertainty, statistical variability, inherent uncertainty, or stochastic uncertainty.

2.3 The measurement **accuracy** describing systematic uncertainty (bias)

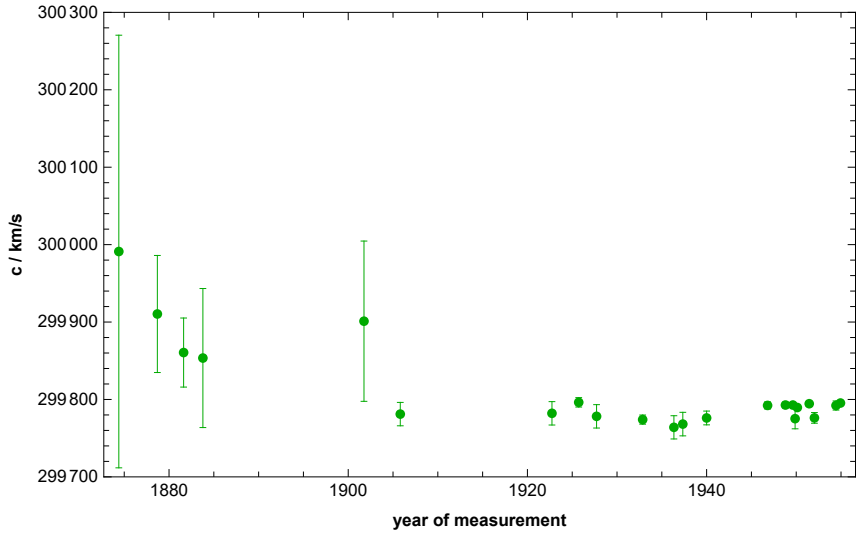


Figure 2: The history of measurements of the speed of light (after [10]).

The measurement of natural constants can be assumed to be undertaken with the highest diligence. This includes even a high level of confidence for the errorbar specifications. In **figure 2** we can see the results for 80 years of measurements of the speed of light. In this area, we can assume the errorbars are not only the simple result of a statistical evaluation but also the result of a carefull examination of all sources of uncertainty. Since the result is assumed to be not influenced by some kind of randomness, the uncertainties are the result of a critical assessment of the experimental setup [11]. The resulting measurement error is called systematic uncertainty or measurement accuracy

$\in_{acc} = 2\,\sigma_{systematic} = f\,(many\ influencing\ factors)$

The most important property of the measurement accuracy is the impossibility to determine it from the measured values by repeating the measurement under constant conditions. It is even impossible to assume a probability distribution function (PDF) within the estimated bounds. This is because the true value is a priori unknown. A posteriori, the only reasonable probability distribution function would be a function with a strong peak at this true value - in principle the PDF is a Dirac delta function for which the position of the peak can be specified only within certain bounds. This problem directly leads to the bounded error approach as introduced by Schweppe in 1968 [12].

For increasing the accuracy of a measurement, repetitions are useless. One has to design and implement a more accurate experiment. The confidence in the specification of a measurement accuracy is usually rather low and one has to rely on an additional safety factor between 2 for measurements traceable to national standards and 8 for measurements undertaken by diligent scientists. In everyday life, the most important systematic error results from the bias of the people undertaking the measurements. Many accuracy specifications only reflect their wishfull thinking or the commercial interests of the measurement equipment suppliers. While time can be measured with an accuracy of 10^{-15}, for engineering quantities even an accuracy of 10% can be extraordinary accurate (see e.g. [13] for the first film boiling HTC measurement with a justified accuracy specification).

The systematic error is sometimes called **epistemic uncertainty** referring to any lack of knowledge from any source, e.g. simplifications, idealizations, numerical approximations or any unanticipated interactions or unknown or untreated phenomena. It is also related to the risk iceberg (see e.g. [14]) as a symbol for the importance of unknown *knowns* and known or unknown *unknowns*.

2.4 The specification of measurement uncertainty and its consequences

Due to the current situation in science and technology, it is wise to classify scientific measurements and their peer reviewed publications into three major categories:

1. **The measurement uncertainty is unspecified**:
   In this case, the measurement and the associated publication can be safely ignored. The optimistic assumption of an uncertainty identical to the last digit of the specification is misleading in most branches of science. Models relying on such data are GIGO.
2. **The measurement uncertainty specification is oversimplified**:
   Often, the measurement uncertainty is provided without any details and justifications. In this case it can be assumed that the given uncertainty is a crude and very optimistic assessment of the precision. When trusting the data, one may use such a specification with an additional trust factor well above 8.
3. **The measurement uncertainty is specified (precision and accuracy)**:
   Depending on the quality of the details one can use such a specification with an additional trust factor of 2 in the case all measurements showing traceability to national and international standards. Critical evidence relying on such measurements should use trust factors above 8.

The minimum requirement on the specification of a measurement uncertainty $\in_{total}$ is to specify the mean or median value $\overline{x}$ and a measure for the 95% confidence interval:

$$X = \overline{x} \pm \epsilon_{total} \quad with \quad \epsilon_{total} = 2\sqrt{\sigma^2_{random} + \sigma^2_{systematic}}$$

In the past, the low level of attention for uncertainty issues implies a high *being unfaked* probability for the share of publications with correctly specified and discussed uncertainty. With rising attention (depending on commercial relevance), the share of faked uncertainty specifications will become larger and these may not be as easily identifiable as in the past. The high numbers of decimal places provided by commercial measurement devices allows no assessment of the uncertainty provided by these devices, one has to rely on traceable calibration and accuracy certificates. To exemplify this fact, one can put all available temperature and humidity meters at the same place and compare their displays.

3. The **model uncertainty**

The modeller has the responsibility to provide and justify quantitative error information for the predictions of his model. Without giving away some of the possible accuracy, the only reliable way to perform this task is to compute the uncertainty information for a specific simulation. This computational task can be performed by the model itself, i.e. a hard task for the modeller, or it can be performed by special library functions able to perform simulations by themselves, i.e. a hard computational task. The design of this computational functions and even their usage requires some fundamental knowledge and some quantitative information on the uncertainties of the model inputs as well as its measureable outputs.

3.1 Modell inputs with errorbars

Model inputs (variables) are usually specified by “exact” numbers and thus the model outputs contain no information on their uncertainty. The simplest possible uncertainty specification is the **bounded error approach** [12]. It allows the propagation of ignorance by specifying the model inputs as numeric intervals and computing intervals for the model outputs.

3.2 Error propagation

For uncorrelated x_i and gaussian distributed errors, the approximate precision ΔF of a quantity $F(x_1, .., x_n)$ can be calculated from the $\triangle x_i$

$$\triangle F \approx \sqrt{\sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i} \triangle x_i \right)^2}$$

This is called the *error propagation law* and it may be used to provide uncertainty information for the special case of uncorrelated model inputs having no systematic error compared to their dominating random error, whichever is assumed to be small and gaussian distributed.

3.3 The error bar approach for a single fit parameter

If the model has only one fit parameter, we can also specify the uncertainty of this fit parameter in the same way by an interval and the prediction uncertainty can be calculated easily. This is the conventional **error bar approach**. With an increasing number of fit parameters it becomes more

and more useless.

If the model has two or more fit parameters, we have to consider the uncertainty of these fit parameters. To do this, we need a valid concept for the specification of fit parameters including their uncertainty and we need to motivate its application by comparing it with the widely used error bar approach. This requires at least a concept representing the simplest possible case of uncertainty specification for model fit parameters relying on the simplest possible case of measurement uncertainty specification, the bounded error approach. From the viewpoint of statistics it is the (quite frequently occuring) case where there is no hope of identifying a single parameter vector and it was probably first mentioned by Frisch in 1934 [15].

3.4 The simplest possible example

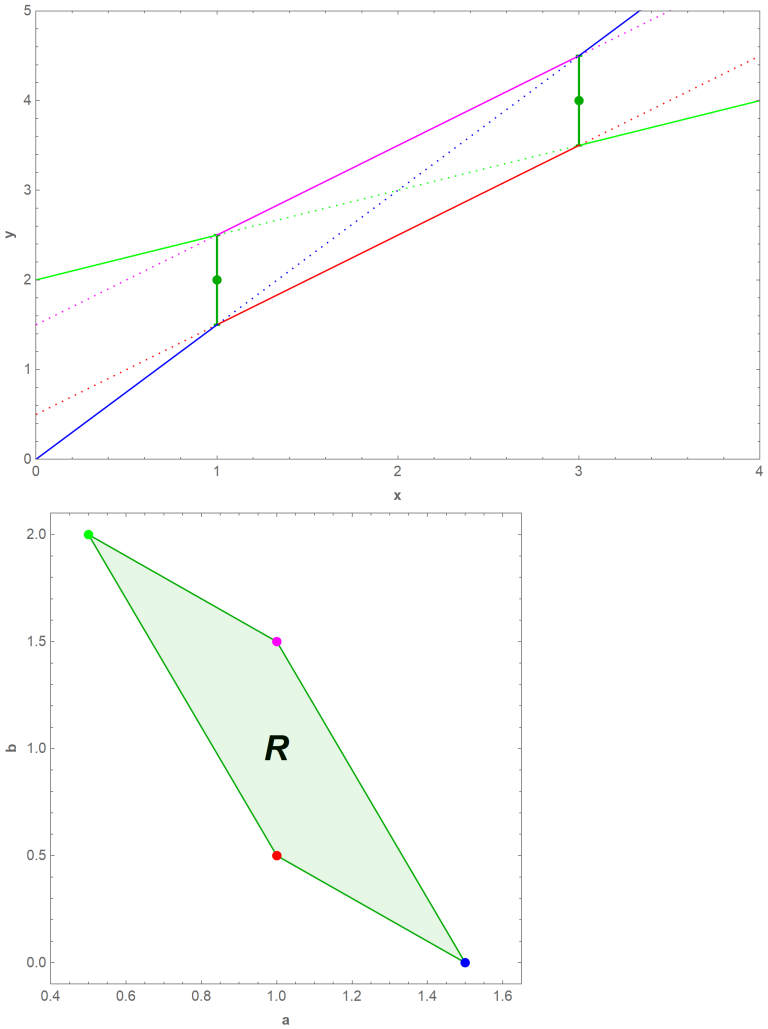


Figure 3. Linear model calibration from 2 measurements with a systematic error.

The simplest possible model with more than one fit parameter is the one dimensional linear model ($n_i = 1,\ n_j = 1,\ n_k = 2$). The model output y depends linearly on the model input x using a slope a and an intercept b

$$y = a\,x + b \tag{1}$$

For the determination of the fit parameters a and b we need a minimum number of n_p=n_k=2 measurements. In the example of **figure 3** (and its n_k-dimensional version) we use $p_k = 1$ for the true

values of the fit parameters. The measured values are not the true values and we want to identify the region of possible true values as a function of the measurement accuracy Δy_p. For the n_k-dimensional generalization of the example, the coordinates of the n_p=n_k measurements are

$$\vec{x}_1 = \vec{1} \text{ and } x_{2..n_p} = \left\{1, \ldots, 1, 3\,\delta_i^{n_p}, 1, \ldots, 1\right\}$$

In the n_k = 2 case from **figure 3**, the accuracy Δy can be calculated for a specific x by calculating the four y_j for the four parameter vectors $\vec{p}_j$

$$\vec{p}_j = \begin{pmatrix} a_j \\ b_j \end{pmatrix} \in \left\{\begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1.5 \\ 0.0 \end{pmatrix}, \begin{pmatrix} 1.0 \\ 1.5 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 2.0 \end{pmatrix}\right\}$$

and the accuracy of the predicted value Δy becomes

$$\triangle y = \frac{\text{Max}\,(y_j) - \text{Min}\,(y_j)}{2}$$

Figure 6 (section 4.2) shows the results for the as simple as possible simple example of **figure 3**.

3.5 The specification of fit parameter uncertainties

The specification of fit parameters by intervals (error bars) is misleading for n_k>1 as the region of possible fit parameters $\mathcal{R}$ is not a hyperrectangle C (cuboid) and, at least for n_k > 5, the volume of such a cuboid C becomes extremely large compared to the volume of $\mathcal{R}$ itself (see the example in table 1 below). Using C instead of $\mathcal{R}$ for calculating the prediction uncertainty, this results in an extreme large uncertainty by and thus a misleading appearance of a very inaccurate model even when the measurements available for the parameter fit support accurate predictions.

Our example - as simple as possible - falsifies the error bar approach for two or more fit parameters and it demonstrates the need to specify at least the **region of possible fit parameters $\mathcal{R}$** in the n_k-dimensional **parameter space $\mathcal{P}$** resulting from a bounded error specification of the models inputs. For the case of dominating random errors this region is the minimum domain of definition for the probability density function $\mathcal{D}(\vec{p})$ of the fit parameters. For the case of relevant systematic errors, the concept of a probability density function in the parameter space is misleading because the *a posteriori* PDF $\mathcal{D}_{\text{true}}(\vec{p})$ becomes more and more peaked at the *a priori* unknown $\vec{p}_{\text{true}}$ and $\mathcal{R}$ is shrinking with decreasing systematic errors.

For the same reason, but to a somewhat lesser extend, it is also misleading to approximate $\mathcal{R}$ by other regular bodies in the parameter space, e.g. an oriented cuboid O or an oriented ellipsoid $\mathcal{E}$ (e.g. [12]). Such a specification may be usefull for small n_k ($\lesssim$ 4) and rapidly becomes useless at higher n_k($\gtrsim$ 5) .

Another way to specify and approximate $\mathcal{R}$ is to identify a representation of $\mathcal{R}$ by a set of n_m parameter vectors $\vec{p}$. Using n_p measurements for identifying the n_k fit parameters n_m is finite ($n_m \leq {n_k}^{n_p}$) for the linear case. We can speak of an **approximation of $\mathcal{R}$ by** these $\hat{\vec{p}}_m$.

As the concept of parameter errorbars is misleading but difficult to eradicate one may think of a less misleading alternative still expressable as a percentage. While model parameters usually have units one has to **normalize all fit parameters first**. This normalization can be done with respect to a typical value of a parameter or a typical span of its values, e.g. from the hyperrectangle spanned by the region $\mathcal{R}$. For the region $\mathcal{R}$ in normalized coordinates, the ratio of the n_k-dimensional volumes ρ=$V_{\mathcal{R}}/V_C$ can be seen as a measure of *parameter accuracy waste* by using a specific C . The

only true and important accuracy is the **accuracy of a specific model prediction** and this can be only **computed from** the **uncertainties of all inputs and an appropriate representation of $\mathcal{R}$** (section 4).

3.6 The simplest (linear) model example in n_k dimensions

n_k	*cuboid*	*or. cuboid*	*ellipsoid*
1	1	n/a	n/a
2	4	1.7	1.6
3	12	3.5	2.7
4	32	7.0	4.9
5	80	15.0	9.2
...	...	...	...
10	5,120	2,783.7	249.0
...	...	...	...
20	10,485,760	n/a	252,020

Table 1. Example from **figure 3** in n_k dimensions: Volume of a geometrical approximation of $\mathcal{R}$ in multiples of the volume of the exact $\mathcal{R}$.

The example from **figure 3** can be formulated in any dimension n_k. As the model is linear, the exact volume of $\mathcal{R}$ as well as the volumes of it's approximations can be calculated. The results shown in table 1 clearly demonstrate **the curse of dimensionality** (a term introduced by R E Bellman in 1961) occuring when trying to approximate $\mathcal{R}$ by geometrical approximations like a cuboid, an oriented cuboid or an ellipsoid. Orders of magnitude in modelling accuracy can be lost by such an approximation.

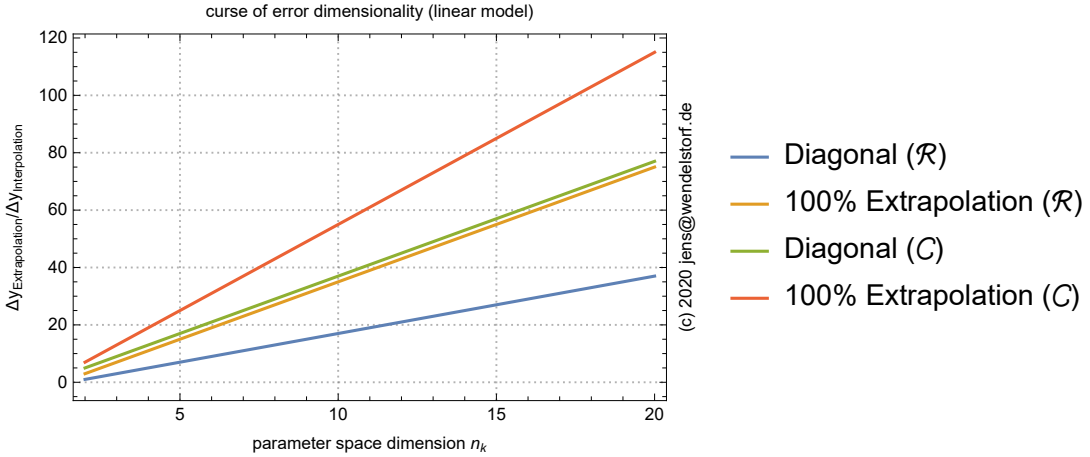


Figure 4. The *curse of error dimensionality*: Error amplification factor for linear model predictions in the extrapolation regime. The model is that of **figure 3** in n_k-dimensions.

As long as we use a linear model for the prediction of results within the direct connections of the measurements used for model calibration, these **interpolations** show the same uncertainty as the underlying calibration measurements (see also **figure 6**). When we try to perform **extrapolations**, the situation changes: extrapolative model predictions show some **curse of error dimensionality**. This is illustrated in **figure 4** for our n_k-dimensional linear model example:

For the example, a minimum number of measurements $\vec{x}_k$ is used and only one factor x_i is varied at a time (still a fairly common approach). From these $n_p = n_k$ measurements and their uncertainties, the region of possible fit parameters $\mathcal{R}$ is computed as well as the cuboid C containing $\mathcal{R}$. C is equivalent to a specification of the p_k with individual uncertainties Δp_k (error bars). In the linear model case, these $\mathcal{R}$ and C can be represented by their 2^{n_k} corners $\vec{p}_m$. Using this representation, the prediction uncertainties can be computed for any model input vector $\vec{x}$. For the $\vec{x}$ lying on some of the connections between the $\vec{x}_k$, the model prediction is an interpolation and it's uncertainty Δy is given by the uncertainty of the measurements $\vec{y}_k$. For all other $\vec{x}$, the model computes an extrapolation and it's uncertainty becomes larger by the **error amplification factor** shown in **figure 4**. Comparing the 4 cases, the smallest factors $\Delta y_{\text{extrapolation}}/\Delta y_{\text{exp}}$ occur for using the exact $\mathcal{R}$ and extrapolating to the $\vec{x}$ in the diagonal of the hypercube spanned by the $\vec{x}_k$ (blue line). In $n_k = 2$ dimensions the diagonal prediction is not an extrapolation and thus the factor is 1. For $n_k = 7$ fit parameters we already loose one order of magnitude in Δy by extrapolating to the diagonal using the exakt region of possible fit parameters $\mathcal{R}$. When we extrapolate further 100% in the diagonal direction, we loose one order of magnitude in Δy for more than 3 fit parameters (orange line - it's proximity to the green line is not typical). When we use the error-bar specification for the parameters (cuboid C), we loose one order of magnitude in Δy for more than 3 fit parameters in the diagonal case (green line) and two orders of magnitude in Δy for more than 16 parameters in the 100% extrapolation case (red line). The slopes of these *curse of error dimensionality*-lines depend on the span of the $\vec{x}_k$ and the proximity of the experimental origin $\vec{x}_0$ to the true origin $\vec{0}$.

For identifiable parameters of a non-linear model any approximation of $\mathcal{R}$ may result in an inability to perform accurate predictions. The specific predictive capability of a non-linear model can depend strongly on the position $\vec{x}$ of the prediction and on the positions of the measurements (and their accuracies as exemplified in **figure 9**) undertaken for the determination of $\mathcal{R}$.

3.7 Statistical viewpoint

The classical statistical viewpoint relies on the least squares approximation (χ^2mimimization [16]) of the parameter vector. Weights are assigned to the measurements. These weights depend on the standard deviations σ of the measurements. The standard deviation of a parameter is then given by the curvature of the χ^2 function at the minimum. Statistical confidence regions in the parameter space can be determined by using a deviation $\Delta\chi^2$ of χ^2 from its minimum depending on the DOF and the target probability (χ^2-distribution). As an approximation one can use the variance-covariance matrix for assessing the reliabilities of the model parameters. Finally, a 95% confidence parameter ellipsoid can be computed.

A convincing procedure to obtain a PDF defined on a region of possible fit parameters and thus some accurate model uncertainty calculation seems to be missing. The main objective is always to get the optimum parameter vector $\vec{p}$ rather than some diffuse approximation of it and the wastefull error bar specification of the fit parameters (confidence intervals) may obviate their application. For an introduction to contemporary statistical methods see [17].

3.8 Conclusions and hypotheses

At this point we may postulate a sufficient understanding of the problem in order to formulate some conclusions and hypotheses:

A lot helps not a lot or *statistics can't solve all problems*

Statistics are always based on the core assumptions of a large number of measurements and some benign probability distribution function. Based on the nature of (the usually dominating) systematic errors and what we have rationalized using our simplest possible example (sub-section 3.4) these assumptions are only valid for a dominating random error ($\in_{\text{prec}} \gg \in_{\text{acc}}$). Conducting more experiments is helpfull only in this case.
Otherwise statistical methods provide no additional benefit as well as conducting more experiments provides only some evidence for the reproducibility declaration. For the indirect measurement of a models region of possible fit parameters $\mathcal{R}$ [18] the position of these measurements (in the $\vec{x}$-space) and their accuracy is decisive. For the case of dominating random errors, statistics should provide a PDF defined on this region.

The measurement accuracy is an important limiting factor

The interpolation capabilities of a model depend on the measurement uncertainties of its inputs Δx_i and of the positions and accuracies of the outputs $y_j \pm \Delta y_j$ used for determining the parameters. For a specific model output y_j, the prediction uncertainty depends on these uncertainties and the corresponding local sensitivities. Even for the case of a linear model, the extrapolation capabilities of a model depend on the extend of extrapolation in the input variables x_i in relation to the $x_{i,\text{range}} = x_{i,\text{max}} - x_{i,\text{min}}$ used for the determination of the parameters. In the case of dominating systematic errors, the mere number of available measurements is not important while their position and accuracy is of critical importance.

Over-parametrization and the *curse of dimensionality*

Without taking the measurement errors into account, it is easy to obtain a perfect fit between model and experiment. Especially for only a few features in the experimental data, the perfect fit can be obtained by increasing the number of fit parameters. By taking into account the measurement errors and using the full region of possible fit parameters, the *curse of dimensionality* uncovers any over-parameterisation. Additionally, fit parameters may be correlated, e.g. if one replaces a by $(a_1 + a_2)$ in equation (1) only the sum is determinable but not the individual values of a_1 and a_2.
The consideration of measurement errors uncovers the problem of structural as well as practical identifiability of a model. It is not necessarily possible to get insight into the internal structure of a system - and thus of its mathematical description - by looking at some input-output measurements [19]. An advantage is the possibility of a computational investigation of every model. Models can be falsified by showing the unidentifiability of their fit parameters and the measurement accuracies required for obtaining a models fit parameters are computeable.

The design of experiment (DOE) problem

The presence of fit parameters requires the determination of an optimal set of experiments to be undertaken for their determination (DOE). In our linear example and for a constant absolute measurement uncertainty there is a simple solution, even in n_k dimensions: One has to undertake an experiment as close as possible to the origin and experiments as far away as possible in all directions defined by the x_i, (not necessarily one direction at a time). All these measurements have to be as accurate as possible in order to get a *small* region of possible fit parameters leading to a maximum prediction accuracy of the model. Already a varying measurement accuracy can change the situation completely and can put the focus on the accuracy of the individual measurements rather than their position. For nonlinear models, the computation of a *least effort* set of experiments for determining the fit parameters can be computationally challenging. It underlines the **iterative structure of any scientific investigation**: The availability of parameterized models allow the prediction of valueable experiments and the available measurements (including their uncertainties) allow the improvement (or falsification) of the models.

The region of possible fit parameters $\mathcal{R}$ is unavoidable and not a simple body

For a model with a single fit parameter, this parameter has to be specified at least by p±Δp. For models having many fit parameters such an errorbar specification becomes more and more counterproductive. For any constructive prediction model error calculation procedure (see next section) at least an approximation of the region of possible fit parameters $\mathcal{R}$ is required. Any approximation by a cuboid (=errorbar specification) or another regular body in the high dimensional parameter space becomes more and more counterproductive with increasing dimensionality of the problem:

- In a high dimensional parameter space, the geometrical shape of $\mathcal{R}$ is often more like a filament than that of a simple and regular body.
- High dimensional parameter spaces are *edge dominated* and thus it makes no sense to look for a specific n_k-dimensional body (e.g. ellipsod) approximating $\mathcal{R}$.
- Only a small number of (accurate) experiments may provide sufficient information for obtaining an approximation of $\mathcal{R}$ by a set of representative parameter vectors.

4. Computing model uncertainties

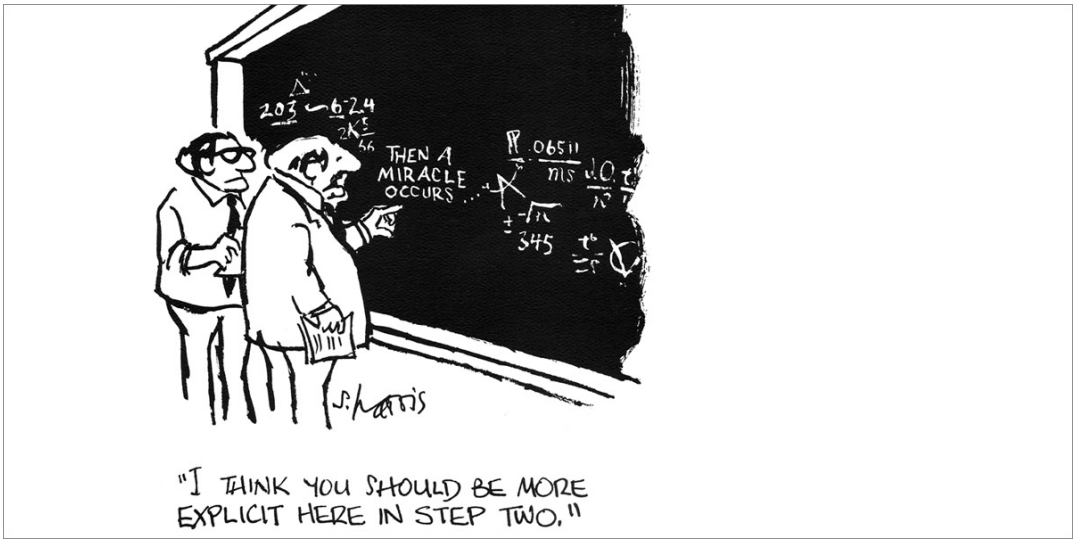


Figure 5. Cartoon from Sidney Harris (1977).

The least effort behaviour in view of the task of providing modelling errors is to ignore them and to state the agreement between the model and some experimental data. This approach is easy to enforce as long as all experimental data is available before the simulations have to be performend and the model has enough fit parameters while the number of features in the data is small.

In day to day modelling, the parameter vector is obtained from a least squares fit, a procedure undertaken for the first time by Gauss and Legendre in 1806 [20]. As long as one does not look at the course of the function minimized during this procedure, one may have some confidence into this parameter vector result. As soon as the model has to predict the results of measurements to be undertaken in the future, a statement on the accuracy of the predictions becomes more and more important. Such a model uncertainty computation should not remind us to the cartoon from Sidney Harris showed in **figure 5**.

But can a least squares fit easily be extended to provide quantitative information on $\mathcal{R}$ or at least some approximation on it? The answer may be hidden in the vast amount of literature that has appeared in that area especially after the extensive review of H L Harter in 1974 [21] or a simple solution of the problem may be still missing but not be missed by those who have to follow the *principle of least effort* [22].

4.1 The bias - free case of pure random error

For the case of dominating random errors (noise), one can use their standard deviations and insert $\in_{\text{prec}}$ into the $\mathcal{R}$ approximation procedure described in the next section. This will provide confidence intervals for the model predictions but no further information, e.g. the most probable value for a model prediction. Statisitical methods [17] are required for such details, but as long as these statistical methods did not provide and use a parameter-PDF at least defined in the region $\mathcal{R}$, their outcomes may remind us to W S Churchills piece of wisdom, “I only believe in statistics that I doctored myself”.

4.2 The bounded error case

As a first step before the application of statistical methods and for the frequent case of dominating systematic errors we need a method for computing the modelling uncertainty for an error in the model inputs specification as simple as possible, but not simpler.
This is the interval representation for the independently measured uncorrelated inputs called **bounded error approach**. This approach can also be used to compute a representation of the region of possible fit parameters $\mathcal{R}$ which is the only lossless concept to account for fit-parameter uncertainties.

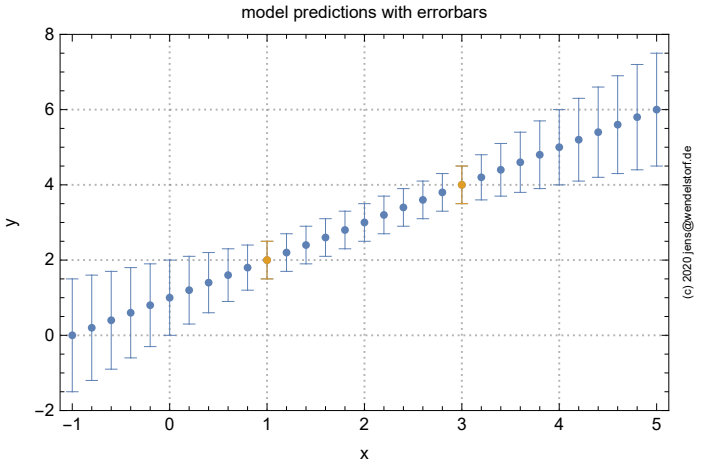


Figure 6. Model prediction for exactly known x - values but model parameters uncertain due to some uncertainty in the measurement of y, the example of **figure 3**.

In **figure 6**, the modelling results for the simple linear model of the preceding section computed for exactly known x-Values are shown. The experiments used for the determination of $\mathcal{R}$ are shown in orange color. In the interpolation regime, their accuracy determines the modelling accuracy (as x is exactly known). In the extrapolation regime, the parameter uncertainties represented by $\mathcal{R}$ lead to a larger uncertainty of the model predictions. As shown in the preceding section, the curse of error dimensionality plays a decisive role for the extrapolative application of models with many fit parameters .

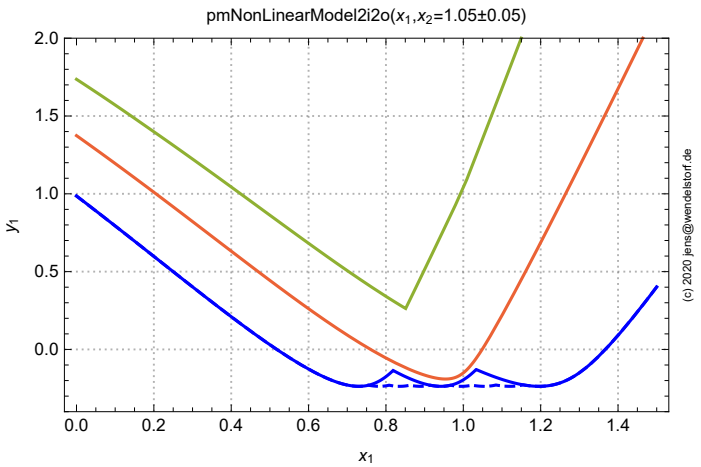


Figure 7. Prediction of a nonlinear model for inaccurately known inputs (inputs ±15%, blue=Min(y_1), red=y_1, green=Max(y_1) - see text and [23].

Nonlinear models may introduce severe sensitivities to both issues, the accurate specification of the inputs as well as the accurate specification of the parameter region $\mathcal{R}$. In **figure 7** the first problem is visualized: The red curve is the normal prediction of the model - as if all model inputs and fit parameters were exactely known. For a better visualization of the prediction uncertainties (green and blue curves), large uncertainties in all inputs (±15%) were choosen. For a linear model, it is sufficient to simulate only the boundaries of all inputs for getting the minimum (blue) and maximum (green) values of a model output. As a compromise one may add the mean values of the inputs to the data base generating the simulation inputs. This means for computing the span of the model outputs we can use *min* and *max* of all inputs (2 samples, not shown because this is sufficient only for the linear case) or a finer sample of values (*n* samples, dashed blue curve is for large *n*). In the case shown in **figure 7**, *n*=3 (blue curve) provides a first estimation of the true result (dashed blue, the maximum - green - curve is not sensitive in this example).

Based on these illustration we can summarize the computational task of providing modelling errors in the next section.

4.3 Computing model uncertainties

A process may include some randomness or some kind of (eventually strange or chaotic) attractors which can not be distinguished from randomness. These phenomena may be seen as some kind of model uncertainty and they can not be neccessarily distinguished from the effects of parameter or other input uncertainties. In this section we will focus on the computation of model uncertainties for any model. And we focus on the calculation of model uncertainties without analysing the internal structure of the model, simply by algorithms performing simulations for specific exact inputs and obtaining exact results for these inputs.

These procedures may also provide some evidence for errors made in the definition phase of the model or the underlying understanding of the process itself. A failure in the parametrization (unidentifiability) or unexpected large uncertainties can be a strong indication for the need to restart the modelling workflow in the definition phase [23].

Even in the case of a perfect model there may be two types of failures unfolded by the procedure of model uncertainty computations:

- The measurements of variables (model inputs) or results (model outputs) required for obtaining the fit parameters may be too inaccurate and the only way to fix it will be to enhance their accuracy, sometimes by orders of magnitude.
- The computational complexity of the approximation of $\mathcal{R}$ may be too large, another kind of unidentifiability.
- The computational complexity of the model uncertainty calculation may be too large.

A failure due to overwhelming computational complexity may have two causes:

1. A large computational complexity of the model itself (the **model computes too long**, e.g. if it includes some CFD simulation). This problem may be fixable by the (possibly automatic) generation of a surrogate model.

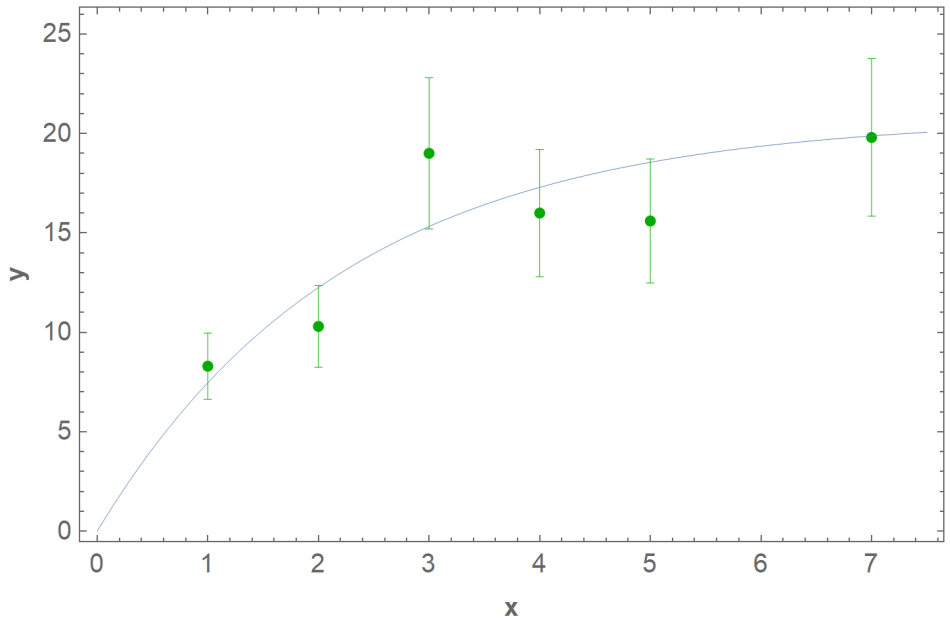
2. A large computational complexity of the uncertainty calculation due to a large number of uncertain model inputs or a haystack of experiments used for parametrization (the **model is called too often**). This problem may be fixable by extracting the key experiments, removing outliers and repetitions and an overall reduction of variable model inputs.

Finally, the computation of model uncertainties consists of the two phases described in the following.

4.3.1 Parameterization phase

In the parameterization phase, we use existing experimental data (measurements) for the quantification of the models fit parameters. In order to take measurement errors into account, we have to accept the following consequences:

- measurement errors (i.e. *error-bars* in the measured inputs and outputs) require a different **concept of fit parameter specification**:
  - For a 1-fit-parameter-model, this single parameter may be specifed by an interval (*error bar for the parameter*).
  - For 2 or more fit-parameters, their specifiction by intervals may result in an unneccessary order of magnitude loss in model accuracy or in unreliable model accuracy computations.
  - For n_k fit parameters, the measurement errors of the model's inputs and outputs have to be considered by (at least approximately) specifiying a region $\mathcal{R}$ (in the n_k-dimensional parameter space) of possible fit parameters.
- the computation of a **set of representative parameter vectors** representing the n_k-dimensional region $\mathcal{R}$ of possible fit parameters is a challenging task depending on the nature of the underlying measurement errors (see sections 4.1 and 4.2 above).
- the cardinality of this set can be the limiting factor leading to the necessity to **reduce the number of fit parameters** in order to be able to compute some approximation of $\mathcal{R}$.
- simple approximations of $\mathcal{R}$, e.g. an n_k-dimensional **ellipsoid**, are **not the panacea** but may be usefull for small n_k<5 [24].
- a large number of fit parameters, i.e. a high dimensional parameter space, is a limiting factor of its own. The **curse of dimensionality** is not only a problem in data science [25,26].



$$Y(x_s, a, b) = a \cdot \left(1 - \mathrm{e}^{-b \cdot x_s}\right)$$

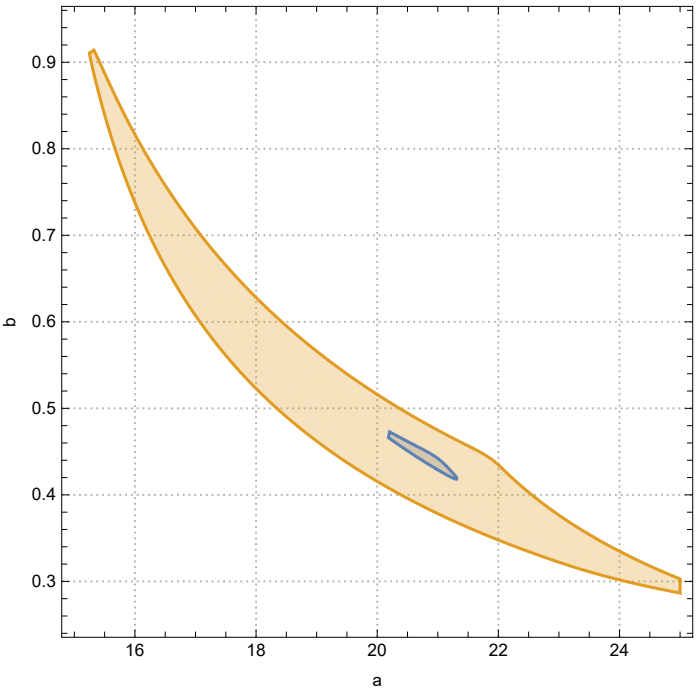


Figure 9. Region of possible fit parameters for a nonlinear example (from [18]) and two different uncertainties of the underlying measurements.

As shown in **figure 9**, $\mathcal{R}$ depends on the magnitude of the measurement uncertainty of the data used for the fit. As this uncertainty is obviously not determineable with a high accuracy of the accuracy specificiation, one should compute $\mathcal{R}$ for some set of factors applied to these accuracies (e.g. 1/2, 1, 2 and 4) and thus study the effect of all possible *real* measurement uncertainties. A vanishing region $\mathcal{R}$ is also an indicator for some outliers in the set of measurements [18], thus data reconciliation is a critical step.

4.3.2 Simulation phase

The naive solution for taking uncertainties of inputs into account is to try to directly calculate with uncertain quantities. This interval calculus [27] was introduced by R E Moore [28]. It induces some mathematical problems and has not found widespread application. It requires changes in the model implementation and thus one may prefer a solution allowing to use conventional models implemented to calculate the outputs from *exact* inputs. Additionally there is an important class of models generating results including some sort of process intrinsic randomness, e.g. a simulation of an epidemic [29]. In this case, a falsifiable modelling result is not the output of a single simulation but formed by the statistical properties of a (sufficiently large) number of simulations. These statistics also influence the parameterization phase discussed in the preceding section.

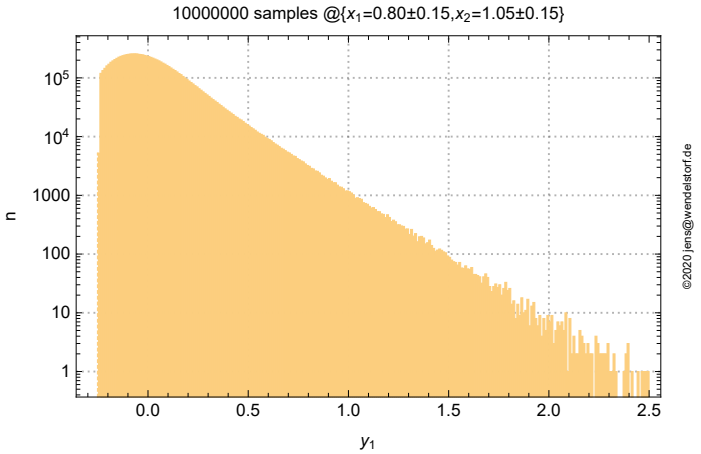


Figure 8. Simulated distribution function [23] of the output of the nonlinear model also used for **figure 7**.

The *gold standard* for studying the effects of uncertainties of the inputs as well as some process inherent randomness on the outputs is called Markov Chain Monte Carlo (MCMC) [30]. It requires a significantly large number of simulations for a representative set of samples representing the uncertainty in the inputs as well as some process inherent randomness. Typical results, as shown in **figure 8**, are distribution functions for model outputs depending on the PDF's of model inputs and (ideally) of the models fit parameters.

4.4 Software support

The majority of modelling and simulation work is undertaken by using existing (commercial) simulation software. Even if it may not be easy to automate the simulation process (the software companies are selling licenses to people) it is allways possible. In the HPC-sector, there are software tools to perform the neccessary sampling, execute the simulations and analyse the results, e.g. EasyVVUQ [31]. The most effective solution is to use a computational language (e.g. the Wolfram Language/Mathematica) for the overall iterative scientific process.

5. Examples from some scientific disciplines

For the limiting case of applicable statistics, the objective of this paper is treated by the statistical **E**rrors **I**n **V**ariables (EIV) Systems methods [32]. Application examples can be found in all areas where it is essential to consider all kinds of uncertainties, e.g. because of the massive consequences of prediction failures (medicine, aerospace, construction engineering, ..) or because of very bad signal-to-noise ratios (exoplanet search).

Uncertainty **Q**uantification (UQ) sees increasing attention in **I**ntegrated **C**omputational **M**aterials **E**ngineering (ICME) [33], computational Materials Science and Simulation-Assisted materials Design [34]. Some specific applications will be mentioned in the following sections.

Thermodynamics (CALPHAD approach)

The equilibrium state of thermodynamic systems can be predicted from the knowledge of the state functions of its phases. This CALPHAD approach introduced by Kaufman in the early 1970ties is based on the availability of complex thermodynamic measurements and their (often unstated) uncertainties. For non-trivial solution phases, the state function (Gibbs-Energy in most cases) is fitted to some thermodynamic model function introducing a significantly large number of fit parameters. The Effects of uncertainties on the resulting models and their predictions were investigated by W B Whiting and his coworkers in the early 1990ties (see [34] and [35]).

The selection and calibration of thermodynamic models may therefore develop towards a combination of precision measurements and computational science, for a first example see e.g. [36].

Materials properties

The prediction of engineering processes requires the indirect measurement of correlated material properties or even plasticity models having more than a few fit parameters. All the issues and tasks of the preceeding sections apply and there are already some attempts to account for these uncertainties [37].

Ecology and earth system science

With increasing complexity of the real natural system to be modelled, the quantitative consideration of simplifications and uncertainties becomes more and more mission critical and are considered and investigated since decades [38,39].

6. Conclusions and outlook

The haystacks of large experimental data sets can be condensed using machine learning methods producing an artifical relation called **machine learning model** (e.g. a neural network) between *inputs* and *outputs* congruent to the measured data set. Extrapolations are not supported and the quality of the artifical relation strongly depends on the feasible choice of the specific method. By now, general purpose relations (laws) can not be extracted automatically as their most important property requires some kind of overview on the sciences as a whole.

With a complete knowledge of the natural laws one may predict the outcome of an experiment even if there is no experimental record in that specific area of system inputs, this is the goal of **scientific modelling**. Some unknown principles or some kind of computational irreducibility may prohibit such an *ab initio* model prediction and one has to rely on some kind of heuristics eventually introducing fit parameters to the model.

Even if the system itself behaves completely deterministic and non chaotic, the uncertainties of the system variables (*inputs*) introduce uncertainties of the system *outputs*. These uncertainties can have different causes and each cause requires some consideration for computing the uncertainties of the systems outputs. If there is no epistemic uncertainty involved, statistical methods allow for the computation of the model uncertainty with the *gold standard* of a **Monte-Carlo simulation**. As an initial estimation or when epistemic uncertainties (section 2.3) dominate, the simplest possible limiting case is the **bounded error approach** where all model inputs and measurements are desribed by an interval of possible values (*error bar*). As a result, there is no longer a best fit, i.e. an unique vector of model fit parameters. Additionally, in the case of multiple fit parameters, there is no interval specification for these individual fit parameters without wasting information contained in the measurements used for the fit. The dimensionality of the parameter space comprises its own **curse of error dimensionality** and the result of the model identification and parameterization procedure should be the combination of a model and some representation of the **region of possible fit parameters** obtained from the available measured *input-output* relations. The shape and volume of this region depends on the model structure and the measurement uncertainties of the system inputs and outputs. The region has a membership function [18]. The volume of the region may be infinite resulting in an unidentifiable model, e.g. for the case of too many or correlated fit parameters.

This fit parameter region representation and the uncertainties of all model inputs allow to compute model predictions including their individual uncertainties. For the case of applicable statistical methods one may even obtain and use a PDF over the parameter region to compute (simulate in most cases) PDF's for the models individual outputs.

The role of uncertainties in modelling and simulation may thus be seen as an issue of additional computational complexity for the parameterization and simulation phase but it does not neccessarily require special models able to compute with uncertainties.

As process models become more and more ubiquitous, they may also show some recursive structure, i.e. a new process model may be formed from a network of existing process models. At least in this case one has to replace wastefull bounded error specifications by a general form of the **region approximation concept**.

Due to the ubiquitous computational resources of our times the day-to-day computation of model uncertainties will become available in more and more scientific disciplines enhancing their trustworthyness and acting as an antidote against modern computer aided *cargo cult science*. Additionally, the costs of the overall procedures, mostly dominated by expenditures for experiments, can be drastically reduced because as soon as some simple model is available, the optimum follow-up experiments can be computed (DOE) and an iterative method of scientific investigation finally provides the model with the highest achieveable predictive power.

References

1. M Baker (2016): 1,500 scientists lift the lid on reproducibility.
Nature **533** 542; DOI: 10.1038/533452a
2. J P Ioannidis et.al. (2017): A manifesto for reproducible science.
Nature Human Behaviour **1** 1; DOI: 10.1038/s41562-016-0021
3. K Popper (1935): Logik der Forschung. In P Frank and M Schlick (eds):
Schriften zur wissenschaftlichen Weltauffassung. *Springer*;
DOI: 10.4324/9780203994627
4. P Bauer, A Thorpe and G Brunet (2015): The quiet revolution of numerical weather prediction.
Nature **525** 47; DOI: 10.1038/nature14956
5. J Wendelstorf (2007): Metallurgical Process Modelling.
In A Ludwig (ed.) *Proc. SteelSim 2007* pp. 433-438 (PDF)
6. J Wendelstorf (2016): Prozessmodellierung in der Hochtemperaturverfahrenstechnik.
Habilitationsschrift, TU-Clausthal; DOI: 10.21268/20160212-111545
7. H G Bock (2018):
Mathematical modelling. Simulation and optimization - a key technology for the 21st century.
Lecture in Suzdal available from 56-hFXmQ7gw@YouTube.
8. L Ljung (1982): Aspects on the system identification problem.
Signal Processing **4** (5-6) 445; DOI: 10.1016/0165-1684(82)90058-5
9. H Zhou (2014): Error propagation. Chapter 15.2 on pages 33-42 in
W F McDonough (ed): Treatise on Geochemistry, Vol. **15**:
Analytical geochemistry - inorganic instrument analysis. *Elsevier*;
DOI: 10.1016/B978-0-08-095975-7.01402-9
10. M Henrion and B Fischhoff (1986): Assessing uncertainty in physical constants.
American Journal of Physics **54** (9) 791; DOI: 10.1119/1.14447
11. M Grabe (2014): Measurement uncertainties. *Springer*;
DOI: 10.1007/978-3-319-04888-8
12. F Schweppe (1968):
Recursive state estimation: Unknown but bounded errors and system inputs.
IEEE Transactions on Automatic Control **13** (1) 22;
DOI: 10.1109/TAC.1968.1098790
13. J Wendelstorf, R Wendelstorf and K H Spitzer (2008):
Spray water cooling heat transfer at high temperatures and liquid mass fluxes.
International Journal of Heat and Mass Transfer **51** 4902;
DOI: 10.1016/j.ijheatmasstransfer.2008.01.032
14. J Wendelstorf (2016): Beiträge der Wissenschaft zur Industrie 4.0!?
Hochschulschrift TU Clausthal; DOI: 10.21268/20161102-111057
15. R Frisch (1934): Statistical confluence analysis by means of complete regression systems.
University Institute of Economics, Publication No. 5/1934, Oslo, Norway.

16. P R Bevington and D K Robinson (2003):
Data reduction and error analysis for the physical sciences (3rd edition). McGraw-Hill

17. B Efron and T Hastie (2016): Computer Age Statistical Inference. Cambridge University Press.
DOI: 10.1017/CBO9781316576533

18. J Wendelstorf (2019): The accuracy of indirect measurements and model calibration by falsification. *Hochschulschrift TU Clausthal,* DOI: 10.21268/20190515-4

19. R Bellman and K J Aström (1970): On structural identifiability.
Mathematical Biosciences **7** (3-4) 329. DOI: 10.1016/0025-5564(70)90132-X

20. S M Stigler (1981): Gauss and the Invention of Least Squares.
The Annals of Statistics **9** 465 (@JSTOR)

21. W L Harter (1974): The Method of Least Squares and Some Alternatives (Part I-VI). *International Statistical Review* **42** (2) 147, **42** (3) 235, **43** (1) 1, **43** (2) 125, **43** (3) 269 & 273, **44** (1) 113;
DOI: 10.2307/1403077, 1402983, 1402658, 1402897, 1403110, 1402966, 1403111

22. G K Zipf (1949): Human behaviour and the principle of least effort.
Addison Wesley, Cambridge, MA.

23. J Wendelstorf (1988-2020): ***pmLib*** A library for scientific process modelling.
see also: J Wendelstorf (2016-): Einführung in die Prozessmodellierung für Ingenieure.
Lectures W7925 & S7903 at TU Clausthal.

24. E Walter and L Pronzato (1997): Identification of parametric models from experimental data.
Springer. (@google)

25. D L Donoho (2000): High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality.
American Math Society on Math Challenges of the 21st century (lecture).

26. J H Friedman: On Bias, Variance, 1/f--Loss, and the Curse-of-Dimensionality.
Data Mining and Knowledge Discovery **1** 55; DOI: 10.1023/A:1009778005914

27. L Jaulin, M Kieffer, O Didrit and E Walter (2001): Applied Interval Analysis. Springer.
DOI: 10.1007/978-1-4471-0249-6

28. R E Moore (1962): Interval arithmetic and automatic error analysis in digital computing.
PhD dissertation, Stanford University

29. J Wendelstorf (2020): Die Berechnung einer Epidemie als computational document.
Hochschulschrift TU Clausthal; DOI: 10.21268/20200921-5

30. Y A Shreider, I M Sobol, N P Buslenko (1966): The Monte Carlo Method: The Method of Statistical Trials. Pergamon; DOI: 10.1016/C2013-0-01870-1

31. R A Richardson et.al. (2020): EasyVVUQ: A Library for Verification, Validation and Uncertainty Quantification in High Performance Computing.
Journal of Open Research Software **8** (1) 11; DOI: 10.5334/jors.303

32. R J Carrol *et.al.* (2006): Measurement error in nonlinear models: a modern perspective.
CRC Boca Raton, 2nd ed.

33. J H Panchal *et.al.* (2013): Key computational modeling issues in Integrated Computational Materials Engineering. *Computer-Aided Design* **45** 4; DOI: 10.1016/j.cad.2012.06.006

34. P Honarmandi and R Arroyave (2020):
Uncertainty Quantification and Propagation in Computational Materials Science and Simulation-Assisted Materials Design. *Integrating Materials and Manufacturing Innovation* **9** 103;
DOI: 10.1007/s40192-020-00168-2

35. M Stan *et.al.*(2019): Uncertainty quantification and propagation in CALPHAD modeling. *Modelling and Simulation in Materials Science and Engineering* **27** 034003; DOI: 10.1088/1361-651X/ab08c3

36. N H Paulson et.al. (2019):
Bayesian strategies for uncertainty quantification of the thermodynamic properties of materials. *International Journal of Engineering Science* **142** 74; DOI: 10.1016/j.ijengsci.2019.05.011

37. B Kouchmeshky and N Zabaras:
The effect of multiple sources of uncertainty on the convex hull of material properties of polycrystals.
Computational Materials Science **47** 342; DOI: 10.1016/j.commatsci.2009.08.010

38. M B Beck et.al. (1981): Uncertainty and arbitrariness in ecosystems modelling: A lake modelling example. *Ecological Modelling* **13** 87; DOI: 10.1016/0304-3800(81)90008-9

39. T F Stocker et.al. (2013): Climate Change 2013: The Physical Science Basis
DOI: 10.1017/CBO9781107415324

Personal remark

The falsification principle is much more constructive than common thought suggests!
The discovery (invention) of the parameter region membership function [18] has provided a lot of insight into the real scientific obstacles to me . Please send any specific comments and publishable applications (data from real measurements + model candidates) which may allow to publish a computable document to the author, especially if you are interested in a joint publication.